制約付最適化問題に対する逐次二次計画法における更新行列のサイジング(数値計算アルゴリズムの現状と展望II)

小笠原, 英穂; 矢部, 博

数理解析研究所講究録
915巻 1995年 203-208

制約付最適化問題に対する逐次二次計画法における更新行列のサイジング

東京理科大学・理学部　小笠原 英穂　(OGASAWARA Hideho)
東京理科大学・工学部　矢部 博　(YABE Hiroshi)

1. はじめに

次のような等号制約付最小化問題

$$\text{(ECP)} \qquad \min \quad f(x), \quad x \in R^n \qquad \text{s.t.} \quad h(x) = 0$$

を考える. ここで $f: R^n \to R^1$ と $h: R^n \to R^m$ $(m < n)$ は $\mathbf{C}^2$ 級と仮定する.

逐次二次計画 (Sequential Quadratic Programming, SQP) 法は, 制約付最小化問題に対して現在知られている数値解法のうちで, 最も有効な方法の1つとされている.

SQP 法のアルゴリズムの基本的な枠組は, 各反復で 1) 探索方向 d_k を QP 部分問題

$$\text{EQP}(x_k, B_k) \qquad \min_d \ \tfrac{1}{2}d^T B_k d + \nabla f(x_k)^T d \quad \text{s.t.} \quad h(x_k) + \nabla h(x_k)^T d = 0$$

の解として求め, 2) 直線探索によりステップ幅 α_k を決定した後, 次の点を $x_{k+1} = x_k + \alpha_k d_k$ で生成して, 3) 行列 B_k を更新する, という手順である.

実際のアルゴリズムでは, α_k を決定するための直線探索評価関数と B_k の更新公式をどのように選択するかが問題となる. 我々は評価関数と更新公式の適合性の観点から, どちらも拡張ラグランジュ関数に基づいて構成された, 従来よりも適合的と考えられる SQP 法のアルゴリズムを提案した [5].

ところで最近, 無制約最小化問題に対する準ニュートン法に対して, 1970 年代に研究された更新行列のサイジングが新たな観点から見直され, その効果について再評価されつつある [3]. SQP 法は無制約問題に対する準ニュートン法を制約付問題に自然に拡張したものと解釈することができる. したがって SQP 法の更新公式にも同様の考えでサイジング手法を取り入れることが考えられるが, 我々の知る限りではこれまでほとんど研究されていないようである.

そこで本稿では, 先に我々の提案した SQP 法のアルゴリズムにサイジングを組み込むことを提案し, 計算効率を改善するためのいくつかの工夫と実際上の留意点について述べる.

2. 無制約最小化問題に対する準ニュートン法

SQP 法において基本となるのは無制約問題に対する準ニュートン法の手法である. したがってまずそれらを概観しておくことにする.

2.1 Broyden 公式族

無制約最小化問題 $\min\{f(x) \mid x \in R^n\}$ に対する準ニュートン法の反復式は

$$x_{k+1} = x_k - \alpha_k B_k^{-1} \nabla f(x_k), \qquad k = 0, 1, \ldots, \tag{1}$$

である. ここで α_k は目的関数 $f(x)$ を直線探索評価関数に用いて決定されるステップ幅である. B_k はヘッセ行列 $\nabla^2 f(x_k)$ の近似行列であり, 新しい近似行列 B_{k+1} はセカント条件

$$B_{k+1}s_k = y_k, \quad \text{ただし} \ \ s_k = x_{k+1} - x_k, \ \ y_k = \nabla f(x_{k+1}) - \nabla f(x_k)$$

を満たすような更新公式によって生成される. 特に DFP 公式と BFGS 公式が有名である.

Broyden はこの両公式を含む 1 パラメタ公式族を提案した.

$$\begin{aligned} B_{k+1} &= \mathrm{Broyden}(s_k, y_k, B_k, \phi_k) \\ &= B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \frac{y_k y_k^T}{y_k^T s_k} + \phi_k (s_k^T B_k s_k) v_k v_k^T. \end{aligned} \tag{2}$$

ただし $v_k = y_k / y_k^T s_k - B_k s_k / s_k^T B_k s_k$, $\phi_k \in R^1$ である. これを B 公式の Broyden 公式族とよぶ. (2) で $\phi_k = 0$ にとると BFGS 公式が得られ, $\phi_k = 1$ とおくと DFP 公式が得られる.

準ニュートン法の大域的収束性を考える上で近似行列 B_k の正定値性の保存は重要である. これについては

$$\phi_k^* = \frac{(y_k^T s_k)^2}{(s_k^T B_k s_k)(y_k^T B_k^{-1} y_k) - (y_k^T s_k)^2} \tag{3}$$

とおいたとき, B_kが正定値で $y_k^T s_k > 0$, $\phi_k > -\phi_k^*$ ならば (2) で生成される B_{k+1} も正定値となることが知られている. Cauchy-Schwarz の不等式から $\phi_k^* > 0$ であることに注意されたい. 通常は直線探索によって $y_k^T s_k > 0$ を満たすような点列が生成されるので, 初期近似行列を正定値にとり, ϕ_k を例えば 0 以上にとっておけば以後の正定値性は保証される.

特に $0 \le \phi_k \le 1$ の場合を convex class という. この class は DFP と BFGS 公式を含むことから, 多くの研究者たちによって関心がもたれ, 種々の理論的・数値的研究がなされている. しかしながら convex class 以外はあまり研究されておらず, その特性はまだよくわかっていないのが現状である.

パラメタ ϕ_k を 1 から 0 へ減少させるにつれて DFP から BFGS へと段々効率がよくなる傾向にあるので, さらに減少させて負の値まで考慮に入れると BFGS よりもさらに効率のよい更新公式が得られるのではないかと期待するのは自然な発想であろう. $-\phi_k^* < \phi_k \le 0$ の場合を preconvex class とよぶ [10]. 最近そのような着眼から preconvex class に対する関心が高まりつつあり, 大域的収束性や局所的超一次収束性などの研究が行われている [10, 1].

2.2 サイジング付 Broyden 公式族とサイジング因子

Oren and Luenberger [8] は Broyden 公式族 (2) にさらにパラメタ $\gamma_k > 0$ を導入し, B_k を γ_k 倍 (サイジング) してから更新する公式族

$$\begin{aligned} B_{k+1} &= \mathrm{Broyden}(s_k, y_k, \gamma_k B_k, \phi_k) \\ &= \gamma_k \Big(B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \phi_k (s_k^T B_k s_k) v_k v_k^T \Big) + \frac{y_k y_k^T}{y_k^T s_k} \end{aligned} \tag{4}$$

を提案した．彼らは $f(x)$ を特に狭義凸2次関数に限定したとき，$Q^{-1}B_{k+1}$ $(Q=\nabla^2 f)$ の条件数が1つ前の条件数よりも小さくなるように γ_k を選ぶことを考え，次式を導出した．

$$\gamma_k^{OL}=\frac{y_k^T s_k}{s_k^T B_k s_k},\qquad \gamma_k^{IOL}=\frac{y_k^T B_k^{-1} y_k}{s_k^T y_k}. \tag{5}$$

(5) の γ_k を順に Oren-Luenberger (OL)，逆 Oren-Luenberger (IOL) サイジング因子とよぶ．

最近 Dennis, Gay and Welsch [4] によって，一般の関数に対しても OL 因子の意味づけがなされた．すなわち OL 因子は $\gamma_k B_k$ の固有値分布が x_k の近くにおけるヘッセ行列 $\nabla^2 f(x)$ の固有値分布と重なるように，Rayleigh 商を利用して決めたものと解釈される．

2.3 逆行列版サイジング付 Broyden 公式族

ヘッセ行列の近似行列 B_k は反復式 (1) からわかるように，逆行列の形で利用される．そこでヘッセ行列の逆行列自体を直接近似行列でおきかえるという考え方もある．

B 公式のサイジング付 Broyden 公式族 (4) において $B_{k+1}^{-1}=H_{k+1}$, $(\gamma_k B_k)^{-1}=\gamma_k^{-1}H_k$ とおくと H 公式とよばれる逆行列版の更新公式族

$$H_{k+1}=\gamma_k^{-1}\left(H_k-\frac{H_k y_k s_k^T+s_k y_k^T H_k}{s_k^T y_k}+\frac{y_k^T H_k y_k}{(s_k^T y_k)^2}s_k s_k^T+(\psi_k-1)(y_k^T H_k y_k)w_k w_k^T\right)+\frac{s_k s_k^T}{s_k^T y_k} \tag{6}$$

が得られる．ただし $w_k=s_k/s_k^T y_k-H_k y_k/y_k^T H_k y_k$, $\psi_k\in R^1$ であり，ϕ_k と ψ_k の関係は

$$\psi_k=\frac{\phi_k^*(1-\phi_k)}{\phi_k+\phi_k^*}\quad\left(=1-\frac{\phi_k(1+\phi_k^*)}{\phi_k+\phi_k^*}\right) \tag{7}$$

である．ここで ϕ_k^* は (3) で B_k に H_k^{-1} を代入したものである．(7) から B 公式の preconvex class $-\phi_k^*<\phi_k\le 0$ は H 公式の $\psi_k\ge 1$ に1対1対応していることがわかる．

3. SQP 法におけるサイジング因子

(ECP) に対するラグランジュ関数 ℓ, 拡張ラグランジュ関数 L はそれぞれ以下の式によって定義される．

$$\begin{aligned}\ell(x,\mu) &= f(x)+\mu^T h(x),\\ L(x,\mu;r) &= f(x)+\mu^T h(x)+\tfrac{1}{2}r h(x)^T h(x)=\ell(x,\mu)+\tfrac{1}{2}r\|h(x)\|^2.\end{aligned}$$

ただし $r>0$ はペナルティパラメタで，$\mu\in R^m$ は等号制約条件に対するラグランジュ乗数である．また $\|\cdot\|$ は ℓ_2-ノルムを表す．

通常の SQP 法では，B_k はラグランジュ関数のヘッセ行列 $\nabla_x^2\ell(x_k,\mu_k)$ の近似とみなされるが，我々は，Tapia (cf. [2]) に従って，拡張ラグランジュ関数のヘッセ行列 $\nabla_x^2 L(x_k,\mu_k;r)$ を近似する行列とみなし，B_k^L で表す．なお μ_k は (ECP) に対するラグランジュ乗数の $k-1$ 反復目の近似ベクトルで，我々のアルゴリズムでは QP 部分問題 EQP(x_{k-1},B_{k-1}^L) のラグランジュ乗数にとっている．

Tapia は解におけるヘッセ行列の構造を考慮し，次の構造化セカント条件を与えた．

$$B_{k+1}^L s_k = y_k^S, \tag{8}$$

$$s_k = x_{k+1}-x_k,\quad y_k^\ell=\nabla_x\ell(x_{k+1},\mu_{k+1})-\nabla_x\ell(x_k,\mu_{k+1}), \tag{9}$$

$$y_k^S = y_k^\ell+r\nabla h(x_{k+1})\nabla h(x_{k+1})^T s_k. \tag{10}$$

B 公式のサイジング付 Broyden 公式族 (4) で y_k, B_k をそれぞれ y_k^S, B_k^L とおけば, Tapia のセカント条件 (8)–(10) を満たす次のようなサイジング付更新公式が得られる.

$$B_{k+1}^L = \text{Broyden}(s_k, y_k^S, \gamma_k B_k^L, \phi_k). \tag{11}$$

これをサイジング付構造化 Broyden 公式族 (SSB) とよぶことにする. この逆行列版である H 公式も (6) から同様にして得られる.

それぞれに対応する OL, IOL 因子 γ_k は (5) から次のようになる.

$$[B\text{ 公式}] \quad \gamma_k^{OL\text{-}S} = \frac{(y_k^S)^T s_k}{s_k^T B_k^L s_k}, \qquad \gamma_k^{IOL\text{-}S} = \frac{(y_k^S)^T (B_k^L)^{-1} y_k^S}{s_k^T y_k^S}, \tag{12}$$

$$[H\text{ 公式}] \quad (\gamma_k^{OL\text{-}S})^{-1} = \frac{s_k^T (H_k^L)^{-1} s_k}{(y_k^S)^T s_k} \left(= \frac{s_k^T B_k^L s_k}{(y_k^S)^T s_k} \right), \qquad (\gamma_k^{IOL\text{-}S})^{-1} = \frac{s_k^T y_k^S}{(y_k^S)^T H_k^L y_k^S}. \tag{13}$$

ここで B 公式の IOL 因子には逆行列が含まれている点に注意する. このため簡単に計算できる場合を除いて, この因子を採用するのは計算量の観点から得策ではない. この因子の使用は, 例えば初期近似行列を単位行列にとって初期サイジングとして用いる場合などに限られることになる.

H 公式の OL 因子にも見かけ上逆行列が含まれているが, QP 部分問題の KKT 条件から

$$B_k^L d_k = -\left(\nabla f(x_k) + \nabla h(x_k)\mu_{k+1}\right) = -\nabla_x \ell(x_k, \mu_{k+1})$$

が成り立つので $B_k^L s_k = \alpha_k B_k^L d_k = -\alpha_k \nabla_x \ell(x_k, \mu_{k+1})$ を利用すれば, 逆行列を回避してこのサイジング因子を計算することができる.

さて無制約問題に対するのと同様に SQP 法においても, preconvex class が BFGS 公式より効率的な更新公式となるか興味のあるところである. しかし B 公式で近似行列の正定値性を保持するためには (3) で与えられる閾値 ϕ_k^* を気にしなくてはならず, 計算量の点からやはり実際的ではない. ところが先述のように B 公式の preconvex class は H 公式の $\psi_k \geq 1$ に対応するので, H 公式によれば閾値を考える必要がない. これも H 公式を用いる際の有利な点といえる.

サイジングには最新情報を用いて少しでもヘッセ行列の近似をよくしてから更新しようという狙いがある. そこでサイジングに用いる y_k^S の近似を高めるためにさらに, 以下の補正を行うことが考えられる. この補正項は解から離れている場合には必ずしも無視できないヘッセ行列の付加的構造を考慮して導出される.

$$y_k^A = y_k^S + r\Big[\nabla h(x_{k+1}) - \nabla h(x_k)\Big] h(x_{k+1}). \tag{14}$$

(12)–(13) で y_k^S の代わりに y_k^A を用いるサイジング因子を $\gamma_k^{OL\text{-}A}, \gamma_k^{IOL\text{-}A}$ と記すことにする. なおサイジング因子だけではなく, セカント条件 (8)–(10) そのものに y_k^A を用いることも考えられよう.

無制約問題に対する準ニュートン法では Shanno and Phua [9] が毎回サイジングを行うより, 初回だけのほうがむしろ全体の計算効率がよいという実験結果を報告して以来, 初期サイジングが一般に広く受け入れられてきた. しかし最近 Contreras and Tapia [3] によっ

て, ある採択基準に従い選択的にサイジングを行ったほうがより効果的であることが実験的に示された. 彼らはまたOL因子やIOL因子とは別のサイジング因子を提案し, BFGS公式を改良する試みも行っている. これらはBroydenのパラメタ ϕ_k をpreconvex classで調整することによってBFGS公式を凌駕しようという試みとともに, いずれもSQP法に取り入れる意義があると思われる.

最後に近似行列の正定値性について注意しておく. 正定値性の保存は無制約問題に対する準ニュートン法と同様SQP法においても重要であり, 特にそれはQP部分問題が一意解をもつための条件にもなる.

正定値性保存のための条件 $y_k^T s_k > 0$ は無制約問題の場合には適当な直線探索法によって達成されるが, 制約付問題の場合, ラグランジュ関数 ℓ は解において必ずしも正定値ではないため, $y_k = y_k^\ell$ に対してこの条件は解の近くでさえ常に満たされるとは限らない. しかし我々は拡張ラグランジュ関数に基づくTapiaの構造化セカント条件(8)–(10)を用いているので $\nabla h(x_{k+1})^T s_k \neq 0$ である限り, 十分大きなペナルティパラメタ r に対して $(y_k^S)^T s_k > 0$ が成立することに注意されたい. ただしセカント条件に y_k^A を用いた場合には必ずしもそうはならないので, アルゴリズムに相応の防止策を組み込んでおく必要がある.

4. おわりに

本稿では制約付最適化問題に対するSQP法にサイジングを組み込むことを提案した. B 公式によるサイジングについては既に我々は[6, 7]の予備的実験から, 一般に効率が悪いとされているDFP公式も適当なサイジングによって十分BFGS公式に匹敵し得るという結果を得ている. これは無制約問題に対する報告[3]と同様の結果である.

今後の課題は前節で述べたいくつかの効率改善の工夫を数値実験によって検証することである. 特に H 公式によるサイジングは B 公式にはない有利な点をもっているので, 検討する価値が十分あると思われる.

参考文献

[1] R. H. Byrd, D. C. Liu and J. Nocedal, On the behavior of Broyden's class of quasi-Newton methods, *SIAM Journal on Optimization* 2 (1992) 533–557.

[2] R. H. Byrd, R. A. Tapia and Y. Zhang, An SQP augmented Lagrangian BFGS algorithm for constrained optimization, *SIAM Journal on Optimization* 2 (1992) 210–241.

[3] M. Contreras and R. A. Tapia, Sizing the BFGS and DFP updates: numerical study, *Journal of Optimization Theory and Applications* 78 (1993) 93–108.

[4] J. E. Dennis,Jr., D. Gay and R. E. Welsch, An adaptive nonlinear least-squres algorithm, *ACM Transactions on Mathematical Software* 7 (1981) 348–368.

[5] 小笠原英穂, 矢部博, 高橋俊彦, 逐次2次計画法における適合性について, 日本数学会秋季応用数学分科会講演アブストラクト, 1993, pp.93–96.

[6] 小笠原英穂, 矢部博, 高橋俊彦, 逐次2次計画法におけるサイジングの効果について, 日本数学会年会応用数学分科会講演アブストラクト, 1994, pp.18–21.

[7] H. Ogasawara and H. Yabe, An application of sizing techniques to the SQP method, *SUT Journal of Mathematics* 30 (1994) 89–106.

[8] S. S. Oren and D. G. Luenberger, Self-scaling variable metric (SSVM) algorithms, Part I: criteria and sufficient conditions for scaling a class of algorithms, *Management Science* 20 (1974) 845–862.

[9] D. F. Shanno and K. -H. Phua, Matrix conditioning and nonlinear optimization, *Mathematical Programming* 14 (1978) 149–160.

[10] Y. Zhang and R. P. Tewarson, Quasi-Newton algorithms with updates from the preconvex part of Broyden's family, *IMA Journal of Numerical Analysis* 8 (1988) 487-509.